

Designing Experiments

Types of empirical studies

- Experiment: you control the intervention
- Observational study: you watch what happens “in the wild”
- Case study: deep dive into examples
- Demonstration: cool story, bro (but not necessarily generalizable)

Many AI “benchmarks” and “studies” are more like demonstrations than experiments even though they use the language of experiments.

Why are we doing this?

What was challenging about the assignment?

Workshopping Experiments

We want to think through potential variations on experiment design.

Some things to consider:

- fit of research question to hypothesis being tested
- confounding variables (is the experiment actual a measure of something else?)
- opportunity for tighter controls
- splitting up experiment
- practical challenges and ways around them

